The Annals of Statistics
2014, Vol. 42, No. 4, 1347–1360
DOI: 10.1214/14-AOS1225

A CHARACTERIZATION OF STRONG ORTHOGONAL ARRAYS OF STRENGTH THREE[1]

BY YUANZHEN HE AND BOXIN TANG

Chinese Academy of Sciences and Simon Fraser University

In an early paper, He and Tang [*Biometrika* **100** (2013) 254–260] introduced and studied a new class of designs, strong orthogonal arrays, for computer experiments, and characterized such arrays through generalized orthogonal arrays. The current paper presents a simple characterization for strong orthogonal arrays of strength three. Besides being simple, this new characterization through a notion of semi-embeddability is more direct and penetrating in terms of revealing the structure of strong orthogonal arrays. Some other results on strong orthogonal arrays of strength three are also obtained along the way, and in particular, two SOA$(54, 5, 27, 3)$'s are constructed.

1. Introduction. Computer models are powerful tools that enable researchers to investigate complex systems from almost every imaginable field of studies in natural sciences, engineering, social sciences and humanities. Computer models can be stochastic or deterministic; we consider deterministic computer models. When the computer code representing a computer model is expensive to run, it is desirable to build a cheaper surrogate model. Computer experiments are concerned with the building of a statistical surrogate model based on the data consisting of a set of carefully selected inputs and the corresponding outputs from running a computer code.

Designing a computer experiment, that is, the selection of inputs, is a crucial step in the process of model building. No matter how elaborate and sophisticated a model building process is, a statistical model contains no more information than what the data can offer. In the past two decades, space-filling designs have been widely accepted as appropriate designs for computer experiments. A space-filling design refers to, in a very broad sense, any design that strews its points in the design region in some uniform fashion. The uniformity of a design may be evaluated using a distance criterion [Johnson, Moore and Ylvisaker (1990)] or a discrepancy criterion [Fang and Mukerjee (2000)]. See Santner, Williams and Notz (2003) and Fang, Li and Sudjianto (2006) for more details. Orthogonality has also been playing a significant role in constructing designs for computer experiments as it, in

Received July 2013; revised December 2013.

[1]Supported by the Natural Sciences and Engineering Research Council of Canada.

MSC2010 subject classifications. Primary 62K15; secondary 05B15.

Key words and phrases. Computer experiment, low dimensional projection, Latin hypercube, space-filling design, (t, m, s)-net.

addition to being useful in its own right, provides a stepping stone to achieving uniformity [Lin, Mukerjee and Tang (2009)].

The curse of dimensionality, however, makes it extremely difficult for the points of a design to provide a good coverage of a high dimensional design region. Even 10,000 points are not enough for a 2^m grid in an $m = 14$ dimensional space, not to mention that such a regular grid leaves a deep hole in the center of the 2^m points. In such situations, it makes more sense to consider designs that are space-filling in lower dimensional projections of the input space. The idea of Latin hypercube designs is to achieve the maximum uniformity in all one-dimensional projections [McKay, Beckman and Conover (1979)]. OA-based Latin hypercubes [Tang (1993)] carry this idea further, which give designs that, in addition to being Latin hypercubes, achieve uniformity in t-dimensional margins when orthogonal arrays of strength t are employed. One could also use orthogonal arrays directly [Owen (1992)] but such designs do not perform well in one-dimensional projections when orthogonal arrays have small numbers of levels.

He and Tang (2013) introduced, constructed and studied a new class of arrays, *strong orthogonal arrays*, for computer experiments. A strong orthogonal array of strength t does as well as a comparable orthogonal array in t-dimensional projections, but the former achieves uniformity on finer grids than the latter in all g-dimensional projections for any $g \leq t-1$. Consequently, Latin hypercubes constructed from a strong orthogonal array of strength t are more space-filling than comparable OA-based Latin hypercubes in all g-dimensional projections for any $2 \leq g \leq t-1$. The concept of strong orthogonal arrays is motivated by the notion of nets from quasi-Monte Carlo methods [Niederreiter (1992)]. The formulation of this new concept has two advantages. First, strong orthogonal arrays are more general than nets in terms of run sizes; and second, strong orthogonal arrays are defined in the form and language that are familiar to design practitioners and researchers. This not only makes existing results from nets more accessible to design community but also allows us to obtain new designs and theoretical results.

The present article focuses on strong orthogonal arrays of strength three. Through a notion of semi-embeddability, we provide a complete yet very simple characterization for such arrays. Though the characterization using generalized orthogonal arrays in He and Tang (2013) is general, our new characterization for strength three is more direct and revealing. Apart from this main result, some other results on strong orthogonal arrays of strength three are also obtained. In particular, we construct two strong orthogonal arrays of 54 runs, five factors, 27 levels and strength three.

The paper is organized as follows. Section 2 introduces some notation and background material. In Section 3, a notion of semi-embeddability is defined, through which we present the main result of the paper, stating that a strong orthogonal array of strength three exists if and only if a semi-embeddable orthogonal array of strength three exists. We then examine the semi-embeddability and nonsemi-embeddability of some orthogonal arrays. Section 4 constructs two SOA(54, 5, 27, 3)'s. A discussion is given in Section 5.

2. Notation and background. This section provides a preparation for the rest of the paper by introducing necessary notation and some background material. An $n \times m$ matrix A with its jth column taking levels $0, 1, \ldots, s_j - 1$ is said to be an orthogonal array of size n, m factors, and strength t if for any $n \times t$ sub-matrix of A, all possible level combinations occur equally often. Such an array is denoted by $\mathrm{OA}(n, m, s_1 \times \cdots \times s_m, t)$ in this paper. If at least two s_j's are unequal, the array is said to be asymmetrical or have mixed levels. When $s_1 = \cdots = s_m = s$, we obtain a symmetrical orthogonal array, in which case, the array is denoted by $\mathrm{OA}(n, m, s, t)$. Since they were first introduced by Rao (1947), orthogonal arrays have been playing a prominent role in both statistical and combinatorial design literature, and have become the backbone of designs for multi-factor experiments. Dey and Mukerjee (1999) discussed the construction and optimality of orthogonal arrays as fractional factorial designs. For a comprehensive treatment of orthogonal arrays, we refer to Hedayat, Sloane and Stufken (1999).

Motivated by the notion of nets from quasi-Monte Carlo methods [Niederreiter (1992)], He and Tang (2013) introduced strong orthogonal arrays. Let $[x]$ denote the largest integer not exceeding x. An $n \times m$ matrix with levels from $\{0, 1, \ldots, s^t - 1\}$ is called a strong orthogonal array of size n, m factors, s^t levels, and strength t if any sub-array of g columns for any g with $1 \le g \le t$ can be collapsed into an $\mathrm{OA}(n, g, s^{u_1} \times s^{u_2} \times \cdots \times s^{u_g}, g)$ for any positive integers $u_1, \ldots, u_g$ with $u_1 + \cdots + u_g = t$, where collapsing into s^{u_j} levels is done by $[a/s^{t-u_j}]$ for $a = 0, 1, \ldots, s^t - 1$. We use $\mathrm{SOA}(n, m, s^t, t)$ to denote such an array. The following is an $\mathrm{SOA}(8, 3, 8, 3)$:

$$\begin{bmatrix} 0 & 0 & 0 \\ 2 & 3 & 6 \\ 3 & 6 & 2 \\ 1 & 5 & 4 \\ 6 & 2 & 3 \\ 4 & 1 & 5 \\ 5 & 4 & 1 \\ 7 & 7 & 7 \end{bmatrix},$$

as we can easily check that:

(i) The array becomes an $\mathrm{OA}(8, 3, 2, 3)$ after the eight levels are collapsed into two levels according to $[a/4] = 0$ for $a = 0, 1, 2, 3$ and $[a/4] = 1$ for $a = 4, 5, 6, 7$.

(ii) Any sub-array of two columns can be collapsed into an $\mathrm{OA}(8, 2, 2 \times 4, 2)$ as well as an $\mathrm{OA}(8, 2, 4 \times 2, 2)$, where collapsing into two levels is done by $[a/4]$ and collapsing into four levels is done using $[a/2]$.

(iii) Any sub-array of one column is an $\mathrm{OA}(8, 1, 8, 1)$.

Lawrence (1996) introduced the concept of a generalized orthogonal array. Extending a result of Lawrence (1996), He and Tang (2013) showed that the existence of a strong orthogonal array is equivalent to the existence of a generalized orthogonal array. For the ease of presentation and the need of this paper, here we give a review of this equivalence result only for the case of strength three. An $n \times (3m)$ matrix $B = \{(a_1, b_1, c_1); \ldots; (a_m, b_m, c_m)\}$ with entries from $\{0, 1, \ldots, s-1\}$, where, as indicated, the $3m$ columns are put into m groups of three columns each, is called a generalized orthogonal array of size n, m constraints, s levels and strength three if all the following matrices are orthogonal arrays of strength three: (a_i, a_j, a_k) for any $1 \le i < j < k \le m$, (a_i, b_i, a_j) for any $1 \le i \ne j \le m$ and (a_i, b_i, c_i) for any $1 \le i \le m$. We use GOA$(n, m, s, 3)$ to denote such an array.

LEMMA 1. *Let* $B = \{(a_1, b_1, c_1); \ldots; (a_m, b_m, c_m)\}$ *be a* GOA$(n, m, s, 3)$. *Define*

$$d_i = a_i s^2 + b_i s + c_i. \tag{1}$$

Then $D = (d_1, \ldots, d_m)$ *is an* SOA$(n, m, s^3, 3)$. *Conversely, if* $D = (d_1, \ldots, d_m)$ *is an* SOA$(n, m, s^3, 3)$, *then* $B = \{(a_1, b_1, c_1); \ldots; (a_m, b_m, c_m)\}$ *is a* GOA$(n, m, s, 3)$, *where* a_i, b_i, c_i *are uniquely determined by* d_i *as given in* (1).

A bit explanation helps understand how a_i, b_i, c_i are obtained from d_i in the second part of Lemma 1. Every integer $0 \le x \le s^3 - 1$ can be uniquely written as $x = x_1 s^2 + x_2 s + x_3$ for some integers x_1, x_2, x_3 with $0 \le x_j \le s-1$. Applying this fact to every component of d_i, we obtain $d_i = a_i s^2 + b_i s + c_i$ for unique vectors a_i, b_i, c_i, all with entries from $\{0, 1, \ldots, s-1\}$.

Strong orthogonal arrays provide a new class of suitable designs for computer experiments. A strong orthogonal array of strength t enjoys better space-filling properties than a comparable orthogonal array in all dimensions lower than t while retaining the space-filling properties of the latter in t dimensions. Strong orthogonal arrays are more general than nets in terms of run sizes. They are defined in the form and language that are familiar to design practitioners and researchers, and thus help to make the existing results from nets more accessible to design community. More importantly, this new formulation of the net idea in terms of orthogonal arrays allows new designs and results to be found, as has been shown in He and Tang (2013) and will be further demonstrated in the next two sections of the present paper.

The rest of the section discusses strong orthogonal arrays in the broad context of quasi-Monte Carlo methods. To approximate an integral, Monte Carlo methods evaluate the integrand at a set of points selected randomly, whereas quasi-Monte Carlo methods do so at a set of points selected in a deterministic fashion. Specifically, to approximate $\int_{[0,1]^m} f(x)\,dx$, quasi-Monte Carlo methods use $\sum_{i=1}^n f(x_i)/n$ where $x_1, \ldots, x_n$ are a set of points in $[0, 1]^m$ that are selected deterministically and judiciously. The Koksma–Hlawka inequality [Niederreiter

(1992), Theorem 2.11] states that

$$\left|\sum_{i=1}^{n} f(x_i)/n - \int_{[0,1]^m} f(x)\,dx\right| \le V(f)D_n^*(P),$$

where $V(f)$ is the bounded variation of f in the sense of Hardy and Krause; $D_n^*(P)$ is the star discrepancy of the set P of points $x_1, \ldots, x_n$, which is defined as the maximum absolute difference between the uniform distribution function and the empirical distribution function based on the point set. According to this result, the set of points for quasi-Monte Carlo methods should therefore be chosen to have a small star discrepancy. When an infinite sequence of points is considered, we use $D_n^*(S)$ to denote the star discrepancy given by the first n points of the sequence. The best general lower bounds on $D_n^*(P)$ and $D_n^*(S)$ are those of Roth (1954) stating that $D_n^*(P) \ge C_m n^{-1}(\log n)^{(m-1)/2}$ for a point set and $D_n^*(S) \ge C_m n^{-1}(\log n)^{m/2}$ for an infinite sequence, where C_m is a constant independent of n. But it is widely believed, though yet to be proved, that

$$D_n^*(P) \ge C_m n^{-1}(\log n)^{m-1}, \qquad D_n^*(S) \ge C_m n^{-1}(\log n)^{m}. \tag{2}$$

Halton sequences and corresponding Hammersley point sets attain the lower bounds in (2), but the implied constants C_m grow superexponentially as $m \to \infty$ [Niederreiter (1992), Chapter 4]. What makes (t, m, s)-nets and (t, s)-sequences attractive is that they have much smaller implied constants while satisfying the lower bounds in (2). Moreover, (t, m, s)-nets and (t, s)-sequences contain an orthogonal array structure, which was pointed out by Owen (1995) and used by Haaland and Qian (2010) to construct nested space-filling designs for multi-fidelity computer experiments.

In what follows, we write (w, k, m)-nets for (t, m, s)-nets and (w, m)-sequences for (t, s)-sequences so as to be consistent in our notation for this paper. An elementary interval in base s is an interval in $[0, 1]^m$ of form

$$E = \prod_{j=1}^{m}\left[\frac{c_j}{s^{d_j}}, \frac{c_j+1}{s^{d_j}}\right),$$

where nonnegative integers c_j and d_j satisfy $0 \le c_j < s^{d_j}$. For $0 \le w \le k$, a (w, k, m)-net in base s is a set of s^k points in $[0, 1]^m$ such that every elementary interval in base s of volume s^{w-k} contains exactly s^w points. Nets and related (w, m)-sequences were first defined by Sobol' (1967) for base $s = 2$ and later by Niederreiter (1987) for general base s.

A deeper connection of nets with orthogonal arrays was established by Lawrence (1996) and independently by Mullen and Schmid (1996). These authors showed that a (w, k, m)-net is equivalent to a generalized orthogonal array. Inspired by this equivalence result, He and Tang (2013) proposed and studied strong orthogonal arrays for computer experiments. Unlike generalized orthogonal arrays, strong orthogonal arrays are in the ready-to-use format and directly capture the space-filling properties of (w, k, m)-nets. The following result is from He and Tang (2013).

LEMMA 2. *If $\lambda = s^w$ for integer w, then the existence of an* SOA$(\lambda s^t, m, s^t, t)$ *is equivalent to that of a (w, k, m)-net in base s where $k = w + t$.*

As strong orthogonal arrays are defined without restricting the index to be a power of s, they provide a more general concept than (w, k, m)-nets. This is in the same spirit as the generalization of orthogonal Latin squares to orthogonal arrays of strength two. He and Tang (2013) discussed several families of strong orthogonal arrays that cannot be obtained from (w, k, m)-nets. Because of Lemma 2, it is not unreasonable to expect that the star discrepancy of strong orthogonal arrays would also be $O(n^{-1}(\log n)^{m-1})$ just like nets, although a precise presentation and rigorous derivation of this result may require some serious work. Since our focus is the finite sample space-filling properties of strong orthogonal arrays, we choose not to dwell any further on the issue of discrepancy in this paper.

3. Characterizing strong orthogonal arrays of strength three. Central to our characterizing result is the notion of embeddability and semi-embeddability for orthogonal arrays.

DEFINITION 1. An orthogonal array OA(n, m, s, t) is said to be embeddable if it can be obtained by deleting one column from an OA$(n, m + 1, s, t)$.

Consider the first column of an OA(n, m, s, t). Then the s levels in this first column divide the whole array into s sub-arrays, which are not orthogonal arrays but all become OA$(n/s, m - 1, s, t - 1)$'s if their first columns are deleted. We say that these s arrays of strength $t - 1$ are obtained by branching the first column. Similarly, branching any other column also produces s orthogonal arrays of strength $t - 1$. In total, ms such arrays of strength $t - 1$ can be obtained. For easy reference, they are called child arrays or simply children of the OA(n, m, s, t) under consideration.

DEFINITION 2. An OA(n, m, s, t) is said to be semi-embeddable if all of its ms children are embeddable.

The following result is immediate.

LEMMA 3. *If an* OA(n, m, s, t) *is embeddable, then it must be semi-embeddable.*

The converse of Lemma 3 is not always true, and we will see many examples in the rest of the paper. One result in He and Tang (2013) states that if an embeddable

$\mathrm{OA}(n, m, s, 3)$ is available, then an $\mathrm{SOA}(n, m, s^3, 3)$ can be constructed. The main result of this paper, the following Theorem 1, provides a complete characterization for the existence of an $\mathrm{SOA}(n, m, s^3, 3)$.

THEOREM 1. *An* $\mathrm{SOA}(n, m, s^3, 3)$ *exists if and only if a semi-embeddable* $\mathrm{OA}(n, m, s, 3)$ *exists.*

The proof, as given in Appendix, is actually constructive, and it shows how to construct an $\mathrm{SOA}(n, m, s^3, 3)$ from a semi-embeddable $\mathrm{OA}(n, m, s, 3)$ and vice versa. While the characterization is fundamental of a strong orthogonal array through a generalized orthogonal array as in He and Tang (2013), Theorem 1 does provide a more direct and penetrating characterization for strong orthogonal arrays of strength three.

As an immediate application of Theorem 1, we present the following result on the maximum number of constraints on strong orthogonal arrays.

THEOREM 2. *We have* $h(n, s, 3) = f(n, s, 3) - 1$, *provided that*

$$f(n, s, 3) = f(n/s, s, 2) + 1, \tag{3}$$

where $h(n, s, t)$ *and* $f(n, s, t)$ *are the largest* m *for an* $\mathrm{SOA}(n, m, s^t, t)$ *and an* $\mathrm{OA}(n, m, s, t)$ *to exist, respectively.*

We know from He and Tang (2013) that $f(n, s, 3) - 1 \le h(n, s, 3) \le f(n, s, 3)$. Theorem 2 then follows from Theorem 1 if we can show that, under the condition in (3), any $\mathrm{OA}(n, m', s, 3)$ with $m' = f(n, s, 3)$ is not semi-embeddable. This is obvious as none of its child arrays, which are $\mathrm{OA}(n/s, m' - 1, s, 2)$'s, can be embeddable due to $m' - 1 = f(n/s, s, 2)$.

For $s = 2$, the condition in (3) is always met, and this special case of Theorem 2 was obtained in He and Tang (2013). Another important case where the condition in (3) holds is when $n = s^3$ and s is an even prime power, in which case we have $f(n, s, 3) = s + 2$ and $f(n/s, s, 2) = s + 1$ [Hedayat, Sloane and Stufken (1999)].

The results of Bierbrauer, Edel and Schmid [(2002), Section 7] can be regarded as a linear version of Theorem 1. As such, the following Propositions 1 and 2 have also been established by these authors albeit in different terminology.

PROPOSITION 1. *A linear orthogonal array* $\mathrm{OA}(s^k, m, s, 3)$ *is semi-embeddable, so long as* $m \le (s^{k-1} - 1)/(s - 1)$.

An orthogonal array $\mathrm{OA}(n, m, s, t)$ is said to be linear if its runs, as vectors based on a finite field $\mathrm{GF}(s)$, form a linear space. Proposition 1 can also be established directly. We omit the details but provide the following pointers for those readers who are interested in a direct proof. Any linear $\mathrm{OA}(s^{k-1}, m - 1, s, 2)$ is a sub-array of the saturated linear $\mathrm{OA}(s^{k-1}, (s^{k-1}-1)/(s-1), s, 2)$ from Rao–Hamming construction [Cheng (2014), Chapter 9]. Permuting the levels within a column of this saturated linear array generates another $\mathrm{OA}(s^{k-1}, (s^{k-1}-1)/(s-1), s, 2)$, which is not linear in general. Any child of a linear $\mathrm{OA}(s^k, m, s, 3)$ is an $\mathrm{OA}(s^{k-1}, m-1, s, 2)$, which is either linear or can be obtained from a linear array by permuting the levels in its columns.

Bush construction gives a linear $\mathrm{OA}(s^3, s+1, s, 3)$, which can be embedded into an $\mathrm{OA}(s^3, s+2, s, 3)$ when s is an even prime power. For odd prime power s, this $\mathrm{OA}(s^3, s+1, s, 3)$ is not embeddable as in this case $f(s^3, s, 3) = s+1$. However, according to Proposition 1, it is semi-embeddable. Therefore, an $\mathrm{SOA}(s^3, s+1, s^3, 3)$ can always be constructed when s is a prime power. Examples are $\mathrm{SOA}(27, 4, 27, 3)$, $\mathrm{SOA}(64, 5, 64, 3)$, $\mathrm{SOA}(125, 6, 125, 3)$, $\mathrm{SOA}(343, 8, 343, 3)$ and so on. We summarize the above discussion in the next result.

PROPOSITION 2. *For any prime power s, we have that $h(s^3, s, 3) = s+1$.*

Consider a linear $\mathrm{OA}(s^4, s^2+1, s, 3)$ based on an ovoid; see Hedayat, Sloane and Stufken [(1999), Section 5.9]. This array satisfies the condition in Proposition 1 and is therefore semi-embeddable. As such, an $\mathrm{SOA}(s^4, s^2+1, s^3, 3)$ can be constructed by Theorem 1. Note that the $\mathrm{OA}(81, 10, 3, 3)$ resulting from taking $s = 3$ is not embeddable as $f(81, 3, 3) = 10$.

If a run occurs more than once in an orthogonal array, it is called a repeated run. The following Theorem 3 asserts that certain orthogonal arrays are not semi-embeddable if they have repeated runs. The proof of Theorem 3 requires the use of a result on orthogonal arrays with repeated runs, and this is presented in Lemma 4.

LEMMA 4. *If there exists an $\mathrm{OA}(2s^t, m, s, t)$ with a repeated run, then we must have $m \le s+t-1$.*

THEOREM 3. *For $s \ge 3$, an $\mathrm{OA}(2s^3, s+2, s, 3)$ containing a repeated run is not semi-embeddable.*

The proofs of Lemma 4 and Theorem 3 are given in Appendix. The bound in Lemma 4 is quite sharp. For example, taking $t = 2$ gives $m \le s+1$, which is attainable by the $\mathrm{OA}(2s^2, s+1, s, 2)$ from juxtaposing two identical $\mathrm{OA}(s^2, s+1, s, 2)$'s where s is a prime power.

4. Construction of SOA(54, 5, 27, 3). In the present section, we discuss the application of the results in Section 3 to the existence and construction of $\mathrm{SOA}(54, 5, 27, 3)$'s. According to Hedayat, Seiden and Stufken (1997), the maximum number m of factors in an orthogonal array $\mathrm{OA}(54, m, 3, 3)$ is five and there are exactly four nonisomorphic $\mathrm{OA}(54, 5, 3, 3)$'s. These four arrays, labeled as I, II, III and IV, are available in explicit form in their paper. To study the existence and construction of SOA(54, 5, 27, 3)'s, Theorem 1 says that it suffices to examine the semi-embeddability of these four nonisomorphic OA(54, 5, 3, 3)'s.

Array I has two repeated runs and array II has one repeated run. By Theorem 3, neither array is semi-embeddable. Thus, no SOA(54, 5, 27, 3) can be constructed from array I or array II. Both arrays III and IV have no repeated run. It is thus possible for them to be semi-embeddable. Our direct computer search shows that this is indeed the case. The two SOA(54, 5, 27, 3)'s constructed from these two arrays using Theorem 1 are given in Tables 1 and 2, respectively. To save space, both of them are presented in transposed forms, with the top half of each table displaying runs 1–27 and the bottom half runs 28–54.

From Proposition 2, we know that $h(s^3, s, 3) = s + 1$ if s is a prime power. Construction of SOA(54, 5, 27, 3) establishes the following result.

THEOREM 4. *We have that* $h(54, 3, 3) = f(54, 3, 3) = 5$.

To gain some insights into the semi-embeddability and nonsemi-embeddability of the four nonisomorphic OA(54, 5, 3, 3)'s, we make use of the enumeration results on orthogonal arrays of 18 runs. Schoen (2009) enumerated all nonisomorphic orthogonal arrays of 18 runs and he found that there are exactly 12 nonisomorphic OA(18, 4, 3, 2)'s, which he labeled as 4.0.i for $i = 1, \dots, 12$ in his paper. Among these 12 arrays, five of them are nonembeddable and the other seven are embeddable. The nonembeddable ones are 4.0.5, 4.0.7, 4.0.10, 4.0.11 and 4.0.12.

For a given OA(54, 5, 3, 3), three child arrays can be obtained by branching each of the five columns. For the first OA(54, 5, 3, 3), array I, among the three child arrays from branching each column, one is isomorphic to 4.0.1 and the other two are isomorphic to 4.0.5. For array II, one of the three arrays from branching each column is isomorphic to 4.0.5 and the other two are isomorphic to 4.0.2. Thus, the early conclusion that arrays I and II are not semi-embeddable can also be drawn from the fact that array 4.0.5 is not embeddable. For array III, all the 12 child arrays from branching columns 1 through 4 are isomorphic to 4.0.4, and the three child arrays from branching column 5 are isomorphic to 4.0.1. For array IV, the 9 child arrays from branching columns 1, 2 and 5 are isomorphic to 4.0.2, and the 6 child arrays from branching columns 3 and 4 are isomorphic to 4.0.4. Since all these child arrays are embeddable, arrays III and IV are semi-embeddable.

5. Discussion and future work. He and Tang [(2013), Theorem 1] presented a general method of constructing strong orthogonal arrays from ordinary orthogonal arrays. For the case of strength three, this result means that an SOA$(n, m, s^3, 3)$ can be constructed from an OA$(n, m+1, s, 3)$. Specifically, let $(a_1, \dots, a_m, a_{m+1})$ be an OA$(n, m + 1, s, 3)$. Then $B = \{(a_1, b_1, c_1); \dots; (a_m, b_m, c_m)\}$ is a GOA$(n, m, s, 3)$ and $D = (d_1, \dots, d_m)$ is an SOA$(n, m, s^3, 3)$, where $(b_1, \dots, b_m) = (a_{m+1}, \dots, a_{m+1})$, $(c_1, \dots, c_m) = (a_2, \dots, a_m, a_1)$, and $d_i = a_i s^2 + b_i s + c_i$

TABLE 1
SOA(54, 5, 27, 3) *constructed from the third* OA(54, 5, 3, 3), *array III*

0	3	0	8	8	18	24	16	3	7	7	9	12	23	6	6	1	5	1	5	9	18	15	21	13	26	4
0	3	1	18	26	6	8	15	5	9	13	6	7	21	7	8	9	18	17	22	0	3	2	4	12	24	13
0	3	10	8	21	5	24	6	20	1	9	7	9	6	13	23	8	4	18	15	2	7	24	15	3	0	17
1	5	10	22	7	25	4	7	20	11	2	11	8	8	12	21	18	15	6	3	24	15	0	6	3	0	25
9	18	9	15	15	14	14	10	18	21	21	25	25	20	0	0	3	6	3	6	7	5	7	5	1	2	12
4	12	15	14	22	11	19	2	2	24	21	25	17	19	11	10	23	26	20	16	13	10	17	14	20	22	25
17	4	5	21	12	19	10	26	22	2	1	15	24	11	20	11	23	25	19	16	14	10	25	23	20	14	16
25	17	19	1	4	12	12	25	14	22	14	2	5	21	18	23	20	10	16	16	26	13	11	22	26	19	11
16	19	16	13	13	1	4	14	26	14	23	23	20	2	5	22	19	10	17	17	26	12	21	9	24	9	18
12	16	16	13	17	13	17	24	24	23	23	26	22	26	22	10	11	11	20	19	19	1	4	4	2	8	8

TABLE 2
SOA(54, 5, 27, 3) *constructed from the fourth* OA(54, 5, 3, 3), *array IV*

0	3	0	8	8	18	24	16	3	7	7	9	12	23	6	6	1	5	1	5	9	18	15	21	13	26	4
0	3	1	18	26	6	8	15	5	9	13	6	7	21	7	8	9	18	17	22	0	3	2	4	12	24	16
0	3	10	8	21	5	24	6	23	1	9	7	9	6	14	19	8	4	18	15	2	7	24	12	3	0	13
1	5	10	22	7	25	4	7	23	11	2	11	8	8	18	12	18	15	6	3	24	12	0	6	3	0	13
9	18	9	15	15	14	14	10	18	21	21	25	25	20	0	0	3	6	3	6	7	5	7	5	1	2	12
4	12	15	14	22	11	19	2	2	21	24	25	17	19	11	10	23	26	20	16	13	10	17	14	25	20	22
14	4	5	21	12	19	10	26	22	2	1	15	24	11	20	11	23	25	19	13	17	10	25	23	14	20	16
26	17	22	1	4	12	15	25	17	20	16	2	5	21	18	20	19	11	13	14	25	16	10	23	22	26	11
25	22	16	13	16	1	4	17	26	20	17	23	20	2	5	19	10	19	14	26	14	15	9	21	9	24	21
12	16	16	13	17	13	17	24	24	23	23	26	22	26	22	10	11	11	20	19	19	1	4	4	8	2	8

for $i = 1, \dots, m$. Note that the same a_{m+1} is taken for all b_i's. This simple construction of strong orthogonal arrays from orthogonal arrays should be sufficient for most practical purposes because only one column is lost during the construction. In the terminology of the present paper, semi-embeddability of $A = (a_1, \dots, a_m)$ is automatic due to its embeddability into $(a_1, \dots, a_m, a_{m+1})$.

The above construction also says that embeddability of an OA$(n, m, s, 3)$ is sufficient for the existence of an SOA$(n, m, s^3, 3)$. The present paper strengthens this result by proving that an SOA$(n, m, s^3, 3)$ exists if and only if a semi-embeddable OA$(n, m, s, 3)$ exists. The path of constructing an SOA$(n, m, s^3, 3)$ from a semi-embeddable OA$(n, m, s, 3)$ is still via a GOA$(n, m, s, 3)$, but there is some difference. To illustrate, let $A = (a_1, \dots, a_m)$ be a semi-embeddable OA$(n, m, s, 3)$. Then in forming $B = \{(a_1, b_1, c_1); \dots; (a_m, b_m, c_m)\}$, a GOA$(n, m, s, 3)$, although we still take $(c_1, \dots, c_m) = (a_2, \dots, a_m, a_1)$, the columns b_i's as obtained in the proof of Theorem 1 in the Appendix cannot be all the same unless the semi-embeddable $A = (a_1, \dots, a_m)$ is also embeddable. This is because if the b_i's equal the same column, say b, then $(a_1, \dots, a_m, b)$ must be an OA$(n, m + 1, s, 3)$ due to the fact that $B = \{(a_1, b, c_1); \dots; (a_m, b, c_m)\}$ is a GOA$(n, m, s, 3)$.

The question arises of if a given orthogonal array is semi-embeddable. The simplest case is to consider sub-arrays of the available orthogonal arrays by deleting one or more columns. All arrays obtained this way are embeddable, and hence semi-embeddable. In Section 3, we have presented two further results for judging whether or not an orthogonal array is semi-embeddable. Proposition 1 tells us that a linear orthogonal array OA$(s^k, m, s, 3)$ is semi-embeddable, provided $m \le (s^{k-1} - 1)/(s - 1)$. This result has led to the conclusion that OA$(s^3, s + 1, s, 3)$ from Bush construction and OA$(s^4, s^2 + 1, s, 3)$ base on an ovoid are both semi-embeddable, where s is any prime power. Theorem 3 states that an OA$(2s^3, s + 2, s, 3)$ is not semi-embeddable for $s \ge 3$ if it has a repeated run, allowing us to immediately identify two nonsemi-embeddable OA(54, 5, 3, 3)'s in Section 4. When none of the above methods can give a definitive answer, one can make use of relevant enumeration results if they are available or conduct a complete search as a last resort. In Section 4, we have done it both ways in determining the semi-embeddability of the other two OA(54, 5, 3, 3)'s.

One obvious future direction is to study to what extent the current work can be extended to strong orthogonal arrays of strength four or higher. Although such extension work may not be as neat as what we have done for strong orthogonal arrays of strength three, some useful results are still possible. We leave this to the future.

A more promising direction is what can be done when orthogonal arrays of strength three or higher are too expensive to use for given resources. As discussed in He and Tang (2013), strong orthogonal arrays of strength two can be straightforwardly constructed from ordinary orthogonal arrays of strength two but the former do not improve upon the latter in terms of lower dimensional space-filling. The question then is if we can construct designs that, although not strong orthogonal arrays of strength three, are better than strong orthogonal arrays of strength two. Some preliminary results have been obtained, and we hope to write a future paper along this direction.

APPENDIX

PROOF OF THEOREM 1. We first prove that the existence of an SOA$(n, m, s^3, 3)$ implies the existence of a semi-embeddable OA$(n, m, s, 3)$. Suppose there exists an SOA$(n, m, s^3, 3)$. Then Lemma 1 implies the existence of a GOA$(n, m, s, 3)$, $B = \{(a_1, b_1, c_1); \ldots; (a_m, b_m, c_m)\}$. We will show that array $A = (a_1, \ldots, a_m)$ is a semi-embeddable OA$(n, m, s, 3)$. That A is an OA$(n, m, s, 3)$ follows directly from the definition of generalized orthogonal arrays. By Definition 2, what remains to be shown is that the children of A are all embeddable. Now consider array $P = (a_1, \ldots, a_m, b_1)$. That $B = \{(a_1, b_1, c_1); \ldots; (a_m, b_m, c_m)\}$ is a $GOA(n, m, s, 3)$ dictates that (a_1, a_j, b_1) is an OA$(n, 3, s, 3)$ for any $j = 2, \ldots, m$. This implies that the array Q obtained by selecting the n/s rows of $(a_2, \ldots, a_m, b_1)$ that correspond to a given level in a_1 must be an OA$(n/s, m, s, 2)$. Clearly, array Q becomes a child of A if the last column is deleted. This shows that all the s children of A from branching column a_1 are embeddable. The same argument also applies to the children from branching other columns of A.

We next show that an SOA$(n, m, s^3, 3)$ can be constructed from a semi-embeddable OA$(n, m, s, 3)$. Suppose that $A = (a_1, \ldots, a_m)$ be a semi-embeddable OA$(n, m, s, 3)$. We will construct a GOA$(n, m, s, 3)$, $B = \{(a_1, b_1, c_1); \ldots; (a_m, b_m, c_m)\}$. Then Lemma 1 allows an SOA$(n, m, s^3, 3)$ to be constructed from B. The last paragraph shows that if (a_1, a_j, b_1) for any $j = 2, \ldots, m$ is an OA$(n, 3, s, 3)$, then all the children of A from branching column a_1 are embeddable. We observe that this argument is entirely reversible, meaning that if all the children of A from branching column a_1 are embeddable, then a column b_1 can be obtained so that (a_1, a_j, b_1) is an OA$(n, 3, s, 3)$ for any $j = 2, \ldots, m$. Similarly, a column b_i for $i = 2, \ldots, m$ can be obtained so that (a_i, a_j, b_i) is an OA$(n, 3, s, 3)$ for any $j = 1, \ldots, i-1, i+1, \ldots, m$. Take $(c_1, \ldots, c_m) = (a_2, \ldots, a_m, a_1)$. Now it is evident that array $B = \{(a_1, b_1, c_1); \ldots; (a_m, b_m, c_m)\}$ is a GOA$(n, m, s, 3)$. $\square$

PROOF OF LEMMA 4. Let c be a repeated run of an OA$(2s^t, m, s, t)$. For $i = 0, 1, \ldots, m$, let n_i be the number of other runs that have exactly i coincidences with c. As c is a repeated run, we must have $n_m \geq 1$. A result from Bose and Bush (1952) states that

$$\sum_{i=j}^{m} \binom{i}{j} n_i = \binom{m}{j} (2s^{t-j} - 1) \qquad \text{where } j = 0, 1, \ldots, t. \tag{4}$$

Choosing $j=t$ in (4) gives $\sum_{i=t}^{m}\binom{i}{t}n_i=\binom{m}{t}$. Combining this equation with the fact that $n_m\geq 1$, we must have $n_t=\cdots=n_{m-1}=0$ and $n_m=1$. Now consider the two equations given by setting $j=t-1$ and $j=t-2$ in (4). Solving these two equations, we obtain $n_{t-1}=2(s-1)\binom{m}{t-1}$ and $n_{t-2}=2(s-1)(s+t-1-m)\binom{m}{t-2}$. As $n_{t-2}\geq 0$, we must have $s+t-1-m\geq 0$, implying that $m\leq s+t-1$. Lemma 4 is proved. □

PROOF OF THEOREM 3. An OA$(2s^3, s+2, s, 3)$ containing a repeated run gives rise to many children that have repeated runs. Let A_0 be any such child array OA$(2s^2, s+1, s, 2)$ with a repeated run. By Definition 2, Theorem 3 will be established if we can show that A_0 is not embeddable. Let c be a repeated run of A_0 and n_i be the number of other runs of A_0 that have exactly i coincidences with c. Applying the results in the proof of Lemma 4 to the case $t=2$ and $m=s+1$, we obtain

$$n_0=n_2=\cdots=n_s=0,\qquad n_1=2(s^2-1),\qquad n_{s+1}=1. \tag{5}$$

We will prove that A_0 is not embeddable by the method of contradiction. Suppose that A_0 is embeddable and let A_0^+ be an OA$(2s^2, s+2, s, 2)$ obtained from A_0 by adding one column. Recall that c is a repeated run of A_0. Let c^+ be the run of A_0^+ corresponding to c. Let n_i^+ be the number of other runs of A_0^+ that have exactly i coincidences with c^+. By Lemma 4, array A_0^+ cannot have a repeated run, implying that $n_{s+2}^+=0$. Noting that A_0 is a sub-array of A_0^+, and combining $n_{s+2}^+=0$ with the results in (5), we obtain $n_0^+=n_3^+=\cdots=n_s^+=0$, $n_{s+1}^+=1$ and

$$n_1^+ + n_2^+ = n_1 = 2(s^2-1). \tag{6}$$

On the other hand, the coincidence equation in (4) becomes

$$\sum_{i=j}^{s+2}\binom{i}{j}n_i^+=\binom{s+2}{j}(2s^{2-j}-1)\qquad\text{where } j=0,1,2. \tag{7}$$

Using the two equations from taking $j=1,2$ in (7) and the already obtained results about n_i^+ for $i=3,\dots,s+2$, we obtain $n_1^+=2s^2-5$ and $n_2^+=s+1$, which gives $n_1^+ + n_2^+ = 2s^2+s-4$. But this contradicts (6) for $s\geq 3$ because $(2s^2+s-4)-2(s^2-1)=s-2\geq 1$ for any $s\geq 3$. The proof is complete. □

REFERENCES

BIERBRAUER, J., EDEL, Y. and SCHMID, W. CH. (2002). Coding-theoretic constructions for (t,m,s)-nets and ordered orthogonal arrays. *J. Combin. Des.* **10** 403–418. MR1932120

BOSE, R. C. and BUSH, K. A. (1952). Orthogonal arrays of strength two and three. *Ann. Math. Stat.* **23** 508–524. MR0051204

CHENG, C. S. (2014). *Theory of Factorial Design*. CRC Press, New York.

DEY, A. and MUKERJEE, R. (1999). *Fractional Factorial Plans*. Wiley, New York. MR1679441

FANG, K.-T., LI, R. and SUDJIANTO, A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall, Boca Raton, FL. MR2223960

FANG, K.-T. and MUKERJEE, R. (2000). A connection between uniformity and aberration in regular fractions of two-level factorials. *Biometrika* **87** 193–198. MR1766839

HAALAND, B. and QIAN, P. Z. G. (2010). An approach to constructing nested space-filling designs for multi-fidelity computer experiments. *Statist. Sinica* **20** 1063–1075. MR2729853

HE, Y. and TANG, B. (2013). Strong orthogonal arrays and associated Latin hypercubes for computer experiments. *Biometrika* **100** 254–260. MR3034340

HEDAYAT, A., SEIDEN, E. and STUFKEN, J. (1997). On the maximal number of factors and the enumeration of 3-symbol orthogonal arrays of strength 3 and index 2. *J. Statist. Plann. Inference* **58** 43–63. MR1437031

HEDAYAT, A. S., SLOANE, N. J. A. and STUFKEN, J. (1999). *Orthogonal Arrays: Theory and Applications*. Springer, New York. MR1693498

JOHNSON, M. E., MOORE, L. M. and YLVISAKER, D. (1990). Minimax and maximin distance designs. *J. Statist. Plann. Inference* **26** 131–148. MR1079258

LAWRENCE, K. M. (1996). A combinatorial characterization of (t, m, s)-nets in base b. *J. Combin. Des.* **4** 275–293. MR1391810

LIN, C. D., MUKERJEE, R. and TANG, B. (2009). Construction of orthogonal and nearly orthogonal Latin hypercubes. *Biometrika* **96** 243–247. MR2482150

MCKAY, M. D., BECKMAN, R. J. and CONOVER, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21** 239–245. MR0533252

MULLEN, G. L. and SCHMID, W. CH. (1996). An equivalence between (t, m, s)-nets and strongly orthogonal hypercubes. *J. Combin. Theory Ser. A* **76** 164–174. MR1406000

NIEDERREITER, H. (1992). Random number generation and quasi-Monte Carlo methods. In *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia.

NIEDERREITER, H. (1987). Point sets and sequences with small discrepancy. *Monatsh. Math.* **104** 273–337. MR0918037

OWEN, A. B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica* **2** 439–452. MR1187952

OWEN, A. B. (1995). Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (Las Vegas, NV*, 1994). *Lecture Notes in Statist.* **106** 299–317. Springer, New York. MR1445791

RAO, C. R. (1947). Factorial experiments derivable from combinatorial arrangements of arrays. *Suppl. J. R. Stat. Soc.* **9** 128–139. MR0022821

ROTH, K. F. (1954). On irregularities of distribution. *Mathematika* **1** 73–79. MR0066435

SANTNER, T. J., WILLIAMS, B. J. and NOTZ, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York. MR2160708

SCHOEN, E. D. (2009). All orthogonal arrays with 18 runs. *Qual. Reliab. Eng. Int.* **25** 467–480.

SOBOL', I. M. (1967). Distribution of points in a cube and approximate evaluation of integrals. *Ž. Vyčisl. Mat. i Mat. Fiz.* **7** 784–802. MR0219238

TANG, B. (1993). Orthogonal array-based Latin hypercubes. *J. Amer. Statist. Assoc.* **88** 1392–1397. MR1245375

ACADEMY OF MATHEMATICS
AND SYSTEMS SCIENCE
CHINESE ACADEMY OF SCIENCES
BEIJING 100190
CHINA
E-MAIL: heyuanzhen@gmail.com

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
SIMON FRASER UNIVERSITY
BURNABY, BRITISH COLUMBIA V5A 1S6
CANADA
E-MAIL: boxint@sfu.ca